A CURVE TRAFFIC SHAPING METHOD FOR VBR MPEG DIGITAL VIDEO SOURCES

BIN ZENG, LU YAO, RUI WANG
1, 2, 3Department of Management Engineering, Naval University of Engineering, China
E-mail: 1zbtrueice@gmail.com, 2luyaox@yahoo.com, 3kingwis@163.com

ABSTRACT

One of the most important traffic types in the future packet-switch networks is high bandwidth, variable-bit-rate (VBR) video. A key challenge of incorporating VBR video traffic into networks with service guarantees lies in the difficulty of finding an appropriate traffic characterization that captures the dynamics of the source. A simple traffic model, such as leaky bucket, will lead to low network utilization, especially for burst traffic. Both multiple leaky bucket and D-BIND traffic models are more accurate traffic models, but they are too complex to implement the policy function for high-speed network. In this study, we shape the traffic sources by using curve traffic specification. And, we derived the closed form formula of call admission control algorithm and show how to find the appropriate $(\sigma, \rho)$ class from the $\sigma$-$\rho$ curve for the traffic sources. We evaluate our method with a set of 10 minute long MPEG-compressed video traces. In the experiments, the performance is over 1.13 times for $\sigma$-$\rho$ curve traffic model to leaky bucket model. And, the network complexity of the proposed scheme is the same as the leaky bucket model.

Keywords: Traffic Shaping Method, Leaky Bucket Algorithm, Variable-bit-rate Video, Call Admission Control

1. INTRODUCTION

Of many traffic classes, delay-sensitive variable bit rate (VBR) video poses a unique challenge. Since the performance of a bounded delay service is largely influenced by three factors: (1) the specification which describes the worst case traffic from a connection, (2) the scheduling discipline the network switches use, and (3) the accuracy of the admission control functions. The main design goal is to maximize the number of connections that can be supported without violating any delay bound guarantees. Many network architectures have been proposed in [1,2,3,4]. In this study, we focus on issue (1) and (3). We improve the network utilization by using burstiness curve traffic specification $\sigma(\rho)$ instead of single $(\sigma, \rho)$, while the complexity of the network still same as the network of single leaky bucket model. We define the burstiness curve $\sigma(\rho)$ of a message as the maximum number of bits that must be buffered at a node if message allocate a fixed rate $\rho$ bps. The network employs DJ regulator to reconstruct the traffic pattern and Early-Due-first (EDF) to schedule the service order of each input packet. By using the call admission formula of [5] which is base on single leaky bucket traffic, we derived the close form admission control formula for burstiness curve traffic model. We focus on the trade-off involved in the selection of the value of $(\sigma, \rho)$ parameters and their impact on the number of connections.

The traffic specification of a connection describes the worst-case traffic that is generated by this connection. Admission control functions use this specification and resource available to determine to either accept or reject a new connection. Since $(\sigma, \rho)$ provide a good tradeoff between accuracy and simplicity; they can be efficiently implemented [6] and they are shown to accurately characterize VBR video in [7]. The aim of this paper is to improve its performance by using burstiness curve as traffic specification and also considering the current loading of the network.

A leaky bucket mechanism has a burst parameter $\sigma$ and a rate parameter $\rho$, and it can be implemented with a single counter and timer. In [8], the tradeoff between the burst $\sigma$ and rate $\rho$ for traffic specification is studied. They expressed the dependency of $\sigma$ on $\rho$ and defined it as a burstiness curve. The allocation rate can be varied between the average rate and peak rate of input data. We show an example of the smallest possible value of $\rho$ for any choice of $\sigma$ as figure 1. Figure 2 is an example of burstiness curve. We can piecewise the burstiness curve and define each corner point $(\sigma_i,
ρ_j) as a class. Which class is best choice for getting maximum number of connections? According to [9], the knee of burstiness curve is good choice. A knee in the burstiness curve has a distinct indicates that for descriptors that are slightly away from the knee, either the σ or the ρ parameter rapidly increases. With our study, this is not always correct. Since the optimal class, depend on the dynamic traffic loading of the network.

2. PROPOSED TRAFFIC SELECTION ALGORITHM

2.1. Class selection problem

In this section, we describe the traffic class selection problem. The preliminaries are as follows:

![Figure 3. The Tandem Network Model](image)

There are N classes of traffic sources pass through a M-switch tandem model, as shown in figure 3. Class j is characterized by A_j(t) = σ_j + ρ_j t. In addition, there is N_j in class j, which guarantees an end-to-end delay, D_j. Assumed that the same set of classes is used for all links on the model. Define d_j^i is the delay of a packet which belongs to j class in i switch. If D_j ≤ D_j, we have d_j ≤ d_j^i, for i=1 to M.

![Figure 4. The Architecture of Switching Node](image)

A rate-controlled scheduler [10], as shown in figure 4, composed of a delay-jitter regulator and Earliest Deadline First (EDF) scheduler, is adopted at each switch. A scheduler implementing the delay-jitter can assure a source that at every switch the input pattern is fully reconstructed as it enters the network and the waiting time at the regulator of each switch does not contributed to the end-to-end delay [11]. The propagation and processing delay are assumed to be zero. Since it is determined by physical and technological constraint. Therefore, the end-to-end delay is the sum of delay queuing in the EDF scheduler at each switch. Based on the results of [12], the extra number of connections of
class \( s \) permitted to enter the network is
\[ N_{s}^{OPT} = \min \{ N_{s}^\sigma, N_{s}^\rho \}, \]
where \( N_{s}^\sigma \) is the extra number of connections depends on \( \sigma \) and \( N_{s}^\rho \) is the extra number of connections depends on \( \rho \).

\[ N_{s}^\sigma = \frac{D_s - \sum_{i=1}^{M} \sum_{k=1}^{l_i-1} N_i (\sigma_k - \rho_i d_k^i)}{\sigma_s \sum_{i=1}^{M} l_i - \sum_{i=1}^{l_i-1} N_i \rho_k} \tag{1} \]

\[ N_{s}^\rho = \min_{i=1}^{M} \left[ \frac{l_i - \sum_{i=1}^{l_i-1} N_i \rho_k}{\rho_i} \right] \tag{2} \]

Where \( l_i \) is the link bandwidth of node \( i \) and \( d_k^i \) are recursively obtained by using the following reductions, for \( k=1 \) to \( s-1 \), \( i=1 \) to \( M \).

\[ d_k^i = \frac{N_i \sigma_1 + \sum_{s=1}^{s-1} N_i (\sigma_k - \rho_i d_k^i)}{l_i - \sum_{s=1}^{l_i-1} N_i \rho_k}, \quad i=1 \text{ to } M \tag{3} \]

\[ D_s = \sum_{i=1}^{M} d_k^i \geq \sum_{i=1}^{M} N_i \sigma_1 + \sum_{s=1}^{s-1} N_i (\sigma_k - \rho_i d_k^i) \tag{4} \]

According to [13], for an input video stream, there is a unique burstiness curve for it. If we match this \( \sigma\rho \) curve with all the \( N \) classes that the network can support. Supposed these classes are \( C_1, C_2, \ldots, C_3 \) and \( C_1<C_2<\ldots<C_3 \). That is, there are \( n \) sets of classes can be used as traffic specification. Which class can get maximum number of connections? Without lose generality, assumed \( \sigma_{C_1} < \sigma_{C_2} < \cdots < \sigma_{C_n} \) and \( \rho_{C_1} > \rho_{C_2} > \cdots > \rho_{C_n} \). Supposed delay requirement is \( D \). How to select optimal class from above \( n \) classes? We will analysis the relation of the admission region of each class with delay requirement in next section.

### 2.2. Optimal class selection formula

From equation (1), we observe \( N_{s}^\sigma \) is inverse proportional to the value of \( \sigma \). Since the \( \sigma \) value of class \( C_i \) is smallest one among all \( C_i \), that \( N_{s}^\sigma \) of class \( C_i \) is maximum. When the delay requirement is stringent, \( N_{s}^\sigma \) is smaller than \( N_{s}^\rho \), that \( N_{s}^{OPT} \) is restricted by \( N_{s}^\sigma \). In this case, we select class \( C_i \) as traffic specification for getting maximum number of connections. Similar, when the delay requirement is large, we select class \( C_n \) as traffic specification. Supposed \( l_{min}=\min \{ l_1, l_2, \ldots, l_M \} \). We analysis the relation of the class selection with delay requirement for tandem network as below:

Case 1: In case of small delay requirement, request class \( C_i \).

When the delay requirement is more stringent, \( N_{s}^\sigma \) is smaller than \( N_{s}^\rho \), the number of connection are restricted by \( N_{s}^\sigma \). Since the \( \sigma \) value of class \( C_i \) is smallest one among all \( \sigma_i \) that the \( N_{C_i}^\sigma \) is the biggest one. In this case, we select class \( C_i \). By following figure 5, we can get a delay requirement \( D_{C_i} \) (Note: there is no \( D_{C_i}^\rho \), see figure 5) such that \( N_{C_i}^\sigma = N_{C_i}^\rho \), i.e.

\[ N_{C_i}^\sigma = \frac{D_{C_i}^\sigma - \sum_{i=1}^{M} \sum_{k=1}^{l_i-1} N_i (\sigma_k - \rho_i d_k^i)}{\sigma_{C_i} \sum_{i=1}^{M} l_i - \sum_{i=1}^{l_i-1} N_i \rho_k} \tag{5} \]

\[ N_{C_i}^\rho = \min_{i=1}^{M} \left[ \frac{l_i - \sum_{i=1}^{l_i-1} N_i \rho_k}{\rho_{C_i}} \right] \tag{6} \]

Case 2: in case of large delay requirement, request class \( C_n \).

When the delay requirement is more loosely, \( N_{s}^\sigma \) is bigger than \( N_{s}^\rho \), the number of connection is restricted by \( N_{s}^\rho \). Since the \( \rho \) value of class \( C_n \) is smallest one among all \( \rho_i \) that class \( C_n \) can support maximum number of connection in this case. Therefore, we select class \( C_n \). When we follow the curve of class \( C_n \) in the figure 4, we get a delay \( D_{C_n}^\rho \), such that \( N_{C_n}^\sigma = N_{C_n}^\rho \).

\[ N_{C_n}^\rho = \frac{D_{C_n}^\rho - \sum_{i=1}^{M} \sum_{k=1}^{l_i-1} N_i (\sigma_k - \rho_i d_k^i)}{\rho_{C_n} \sum_{i=1}^{M} l_i - \sum_{i=1}^{l_i-1} N_i \rho_k} \tag{7} \]
Case 3: For delay require between $D_{c_{k-1}}^*$ and $D_{c_{k}}^*$, select class $C_k$, $1<k<n$.

For a class $C_k$, we can get a delay $D_{c_k}^*$ such that $N_{c_k}^\sigma = N_{c_k}^\rho$, i.e.

$$N_{c_k}^\sigma = \frac{(D_{c_k}^* - \sum_{j=1}^{l} \frac{N_j (\sigma_j - \rho_j d_j^i)}{f - \sum_{j=1}^{l} N_j \rho_j})}{\sigma_k \sum_{j=1}^{l} \frac{1}{f - \sum_{j=1}^{l} N_j \rho_j}}$$

$$= \frac{\sum_{j=1}^{l} N_j (\sigma_j - \rho_j d_j^i)}{\rho_{c_k}}$$

$$N_{c_k}^\rho = \frac{(D_{c_k}^* - \sum_{j=1}^{l} \frac{N_j (\sigma_j - \rho_j d_j^i)}{f - \sum_{j=1}^{l} N_j \rho_j})}{\sigma_k \sum_{j=1}^{l} \frac{1}{f - \sum_{j=1}^{l} N_j \rho_j}}$$

$$= \frac{\sum_{j=1}^{l} N_j (\sigma_j - \rho_j d_j^i)}{\rho_{c_k}}$$

$$D_{c_k}^* = \frac{(\sum_{j=1}^{l} \frac{1}{f - \sum_{j=1}^{l} N_j \rho_j})(x - \sum_{j=1}^{l} N_j \rho_j)}{\rho_{c_k}}$$

$$+ \sum_{j=1}^{l} \frac{N_j (\sigma_j - \rho_j d_j^i)}{f - \sum_{j=1}^{l} N_j \rho_j}$$

3. NUMBER RESULTS

In order to evaluate the efficiency and optimality of our proposed method, a series of numerical examples are presented in this section. Two performance measure indexes are adopted to decide the efficiency and optimality of the joint selection method. One is the network utilization, in terms of the number of admitted connections, which is used to represent the efficiency of the proposed method. Moreover, another index is average number of connections to evaluate the optimality of the selection method.

We define the deterministic multiplexing gain (DMG) as the gain in number of connections (in average) above a peak-rate-allocation scheme that is achieved. The DMG is used to further quantify the improvements of our method. We have three strategies to select class, (1) By our method, we get the optimal performance as the best case shown. Otherwise, in lack the information of the call admission policy, user only can select one of following two approaches, (2) fix class: in this paper, use knee point of $\sigma-\rho$ curve, and (3) peak rate class: use the peak rate as the allocation rate ($\rho$). In this paper, two experiments are employed to evaluate the efficiency and optimality of the proposed method as follows. (1) With five hops tandem network, no any connection initially, and then evaluate the two performance indexes with delay for each new class. (2) With cross network, there have connections initially, and then evaluate.
two performance indexes with delay for each new class.

From [14, 15], we get a 10-minutes segment of 'compressed video traces named 'news' is shown in figure 6. The format, resolution, and frame rate are IBBPBBPBBPBB, 384x288, and 25 frame/sec, respectively. By analyzing the packet stream of input video, we get a $\sigma - \rho$ curve and match this $\sigma - \rho$ curve with the support classes of the network and get 7 sets of ($\sigma$, $\rho$) class which can be used for traffic specification as table 1.

<table>
<thead>
<tr>
<th>Class</th>
<th>$\sigma$ (bit/s)</th>
<th>$P$ (Kbit/sec)</th>
<th>Class delay requirement ($D_i$) for</th>
<th>Class delay requirement ($D_j$) for</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>194</td>
<td>2000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>217</td>
<td>1900</td>
<td>0.145</td>
<td>0.124</td>
</tr>
<tr>
<td>3</td>
<td>245</td>
<td>1800</td>
<td>0.172</td>
<td>0.150</td>
</tr>
<tr>
<td>4</td>
<td>290</td>
<td>1700</td>
<td>0.215</td>
<td>0.185</td>
</tr>
<tr>
<td>5</td>
<td>446</td>
<td>1600</td>
<td>0.350</td>
<td>0.307</td>
</tr>
<tr>
<td>6</td>
<td>566</td>
<td>1500</td>
<td>0.472</td>
<td>0.419</td>
</tr>
<tr>
<td>7</td>
<td>743</td>
<td>1400</td>
<td>0.660</td>
<td>0.572</td>
</tr>
</tbody>
</table>

8.1. Experiment 1: Five hop tandem network with no any connection initially

The network structure is shown in figure 7. The bandwidth between each node was 1000Mbits/sec except for $l_3$, which is 100Mbits/sec. The delay bound of each class that we assumed is listed in table 1.

![Figure 7. The Tandem Network](image)

We focus on the admission region analysis between 1ms to 1.0sec to cover some of the above time points. By equation (1) and (2), the relations of the number of connections of each class with delay requirement are shown in figure 8. We see all curves have two sections, one is linear increase section and the other is constant section. For example, the linear section of class 1 is from delay 0 sec to 0.19 sec, and the constant section is from delay 0.19 sec to infinite. From figure 8, we understand the class 7 is suitable for large delay requirement and the class 1 is good for smaller delay requirement. The relations of delay requirement and optimal class selection are listed in table 2. The performances of three approaches are shown in figure 9. The DMG was 1.30 for our method and was 1.13 for knee point approach. The performance is over 1.15 times for our method to knee point approach in experiment 1.

![Figure 8. The Admission Region Of Class 1 To 7 In Experiment 1](image)

**Table 2. Optimal Class Selection For Experiment 1**

<table>
<thead>
<tr>
<th>Delay Requirement (D)</th>
<th>Optimal class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above 0.66sec</td>
<td>7</td>
</tr>
<tr>
<td>0.472sec~0.66sec</td>
<td>6</td>
</tr>
<tr>
<td>0.350sec~0.472sec</td>
<td>5</td>
</tr>
<tr>
<td>0.215sec~0.350sec</td>
<td>4</td>
</tr>
<tr>
<td>0.172sec~0.215sec</td>
<td>3</td>
</tr>
<tr>
<td>0.145sec~0.172sec</td>
<td>2</td>
</tr>
<tr>
<td>Below 0.145sec</td>
<td>1</td>
</tr>
</tbody>
</table>
8.2. Experiment 2: Cross network

Cross network as shown in figure 10. The bandwidths between node 1 and node 2, node 2 and node 3, node 3 and node 4, node 2 and node 5, node 3 and node 6, were $l_1=100\text{Mbits/Sec}$, $l_2=1000\text{Mbits/Sec}$, and $l_3=100\text{Mbits/Sec}$, $l_4=100\text{Mbits/Sec}$, $l_5=100\text{Mbits/Sec}$ respectively. Route 1 is going through node 1, node 2, node 3, and node 4. Route 2 is going through node 5, node 2 and node 6. The delay bound of each class that we assumed are listed in table 1. The number of current connections is shown in table 2. The number of connections of each class (class 1 to 7) is shown in figure 11. The node 2 needs to consider all the connections come from both route 1 and 2. The performances of three approaches are shown in figure 12. The DMG was 1.17 for our method and was 1.06 for knee point approach. The performance is over 1.11 times for our method to knee point approach in experiment 2.

![Figure 10. The Crossing Network](image)

Table 3. Current No. Of Connections For EXP2

<table>
<thead>
<tr>
<th>Class</th>
<th>End-to-end Delay (ms)</th>
<th>Burst (bits)</th>
<th>Rate (Kbit/sec)</th>
<th>Current conns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low delay</td>
<td>8</td>
<td>500</td>
<td>800</td>
<td>64.</td>
</tr>
<tr>
<td>Medium delay</td>
<td>9</td>
<td>750</td>
<td>80000</td>
<td>500.</td>
</tr>
</tbody>
</table>

![Figure 11. The Admission Region Of Class 1 To 7](image)

![Figure 12. Comparison Of Admission Region Of Different Strategy](image)

4. CONCLUSION

The traffic characterization used for VBR video connections has a significant impact on the number of connections that can be established in a packet-switch network with deterministic service. In this paper, we derive the close form formula to select the optimal class from burstiness curve and also consider the delay requirement and current network loading. Since the computation is simple that it is suitable to be implemented in the high-speed network. The average performance is over 1.13
times for burstiness curve model to single leaky bucket model in our experiments. Since our experiments are base on one MPEG video trace. Currently, we are going to simulate connections with more diverse traffic characteristics and performance requirements to further explore the application of this method on the real network.

REFERENCES:


[7] Chao Luo, Li Yu, “Performance bounds analysis on different scheduling algorithms for networks with self-similar traffic”, Proceeding(s) of the 1st IEEE International Conference on Communications in China (ICCC), Beijing, China, August 15-18, 2012, pp. 103 - 108.


[14] Park Shihyon, J.DeDourek, “Quality of service (QoS) for video transmission”, Proceeding(s) of First International Conference on Ubiquitous and Future Networks, The University of Hong Kong, Hong Kong, China, Jun 7- 9, 2009, pp.142-147.